

PROPOSTA DE DISSERTAÇÃO

Por Sergio Serra

Tema: Gestão transparente de grandes volumes de dados e metadados científicos em repositórios digitais institucionais

Área de Aplicação: Universidades, Institutos de Pesquisa, Bibliotecas, Instituições de Gestão da Memória

Contextualização da temática: Com o objetivo de contribuir para a ampliação do acesso e compartilhamento e preservação do conhecimento científico em diversas áreas do conhecimento, instituições de fomento e de pesquisa investem cada vez mais na gestão transparente de grandes volumes de dados científicos. Proveniência de dados e princípios FAIR (um acrônimo para *Findable, Accessible, Interoperable e Reusable*) estão fortemente presentes nas discussões e práticas contemporâneas da ciência de dados; os princípios tiveram sua aplicação consolidada em 2017 pela Comissão Europeia que passou a exigir a adoção de plano de gestão de dados, com base nesses princípios, por projetos de pesquisa (em especial os transdisciplinares e de *Open Science*) financiados por seus recursos. Desde então, tais princípios passaram a ser forças norteadoras da descoberta, do acesso, da interoperabilidade, do compartilhamento e da reutilização dos dados e metadados de científicos em todo o mundo. Estudos realizados na EUA e Europa comprovam o impacto e o potencial dos metadados de proveniência e dos princípios dando destaque às necessidades e aos exemplos de aplicação em diferentes áreas interdisciplinares. No entanto, devido a ser uma novidade, quando colocados em prática levantam dúvidas e imprecisões, gerando diferentes interpretações, o que dificulta sua aplicação imediata.

Objetivo: O objetivo dessa dissertação é estudar um caso de uso considerando um repositório digital institucional e aplicar os princípios FAIR, estabelecer métricas próprias que medem o nível de FAIRness dos objetos digitais ali dispostos; contribuir com a disseminação de um ecossistema de dados FAIR (por exemplo tecnologias Data FAIRPort, FAIR Data Point e PROV) e padrões de metadados nos repositórios na instituição com vistas a possíveis futuras integrações de repositórios digitais com outras instituições.

Contribuição estimada: Apropriação do conhecimento com aplicação no contexto dos repositórios institucionais da UFRRJ, contribuindo para que alinhe com a rede GO-FAIR; produção de um artigo científico e formação de recursos humanos na área.

Metodologia: 1) Revisão Sistemática da Literatura; 2) Escolha do Repositório; 3) Delineamento do Caso de Uso; 4) Definição de métricas; 5) Implementação e Avaliação; 6) Tabulação e análise de resultados; 7) Apresentação. Os experimentos poderão ser realizados no Laboratório de Banco de Dados no PAP/UFRRJ e no CEDIM.

Cronograma: 2-3 meses revisão; 1-2 meses escolha/estudo do repositório; 2-3 meses delineamento do caso de uso; 3-6 meses definição de métricas e implantação da tecnologia; 2-3 meses experimentação; 2-3 meses de elaboração da dissertação = Total 18 meses

Referências:

<https://www.go-fair.org/fair-principles/>

<https://www.w3.org/TR/prov-overview/>

<https://github.com/FAIRDataTeam>

Perfil de possíveis orientados: Ciência da Informação/Computação com conhecimento básico de língua inglesa